

A COMPARISON OF THE TEACHING EFFECTIVENESS OF TENURED AND
NONTENURED FACULTY AS MEASURED BY STUDENT EVALUATIONS

By

LILLY M. LORETTA JAMES

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1975

Copyright

by

Lily M. Loretta James

1975

To Mutter (Mother)

Who Instilled in Me
The Desire to Know
And Gave Me
The Courage to Do

ACKNOWLEDGMENTS

The writer is deeply grateful to Dr. Kern Alexander, Chairman of her Supervisory Committee, for his constant guidance, encouragement, understanding, and cooperation during the preparation of this disquisition. The writer is also grateful to the other members of her Committee: to Dr. James L. Wattenbarger for his helpful suggestions and his confidence in her and to Dr. Athol B. Packer for the support and cooperation that enabled her to pursue her goal.

A special word of appreciation is extended to Dr. T. Winston Cole, whose resourcefulness and concern made it possible for the actual work on this dissertation to get under way, and to Dr. Vynce Hines, who helped the writer establish and interpret the statistical procedures used in this study.

The writer is indebted to Ms. Janice Moxley for setting up her computer programs and assisting her in the analysis of data and to Ms. Lynn Little for her aid in the gathering of data.

Finally, the writer expresses her gratitude to her dearest friend, John H. Hopkins, Sr., for his continual support and encouragement and his unwavering faith in her; to her mother, Isabella, and her sisters and brothers for their moral support; and to her friend Ann M. Whitehead for spurring her on daily toward the accomplishment of her goal.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
LIST OF TABLES	vii
ABSTRACT	viii
 CHAPTER	
I. INTRODUCTION	1
The Problem	4
Statement of the Problem	4
Delimitations and Limitations	4
Justification for the Study	5
Assumptions	5
Definition of Terms	5
Procedures	6
Study Participants	6
Sources of Data	7
Data Collection	8
Data Treatment	8
Organization of the Study	9
II. BACKGROUND OF THE PROBLEM	10
General Guidelines for Tenure-Granting Institutions	13
Criteria for Determining Tenure Status at the University of Florida	14
III. REVIEW OF RELATED LITERATURE AND RESEARCH	19
Tenure	19
The Tenure Debate	22
Tenure and Accountability	24
Teaching Effectiveness	25
Teaching Effectiveness and Accountability	27
Teaching Effectiveness and Tenure	30
Teaching Effectiveness as Defined by the Public in the Mid-1970s	31
Student Evaluation as a Measure of Teaching Effectiveness	33
Student Evaluations	34
Use of Student Evaluations	35

CHAPTER	Page
Faculty Views Regarding Student Evaluations	38
Students' Concepts of the Effective Teacher	40
Reliability of Student Evaluations	42
Validity of Student Evaluations	44
Inconclusive Evidence Regarding Student Evaluations	49
IV. PRESENTATION OF DATA	51
Statistical Analysis	51
Multivariate Analysis of Variance	52
Discriminant Analysis	55
Summary Measures in the Treatment of Data	58
V. ANALYSIS AND DISCUSSION OF DATA	65
Faculty Members Classified by Group	65
Discriminant Analysis on Ten Measures	66
Comparison of Student Ratings for Tenured and Nontenured Faculty	69
Comparison of Student Ratings for Professors and Associate Professors	70
Comparison of Student Ratings for Associate Professors and Assistant Professors	72
Summary Statement	72
VI. FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS	74
Findings	74
Conclusions	77
Recommendations	77
BIBLIOGRAPHY	79
BIOGRAPHICAL SKETCH	84

LIST OF TABLES

1.	Faculty Members Classified by Group	52
2.	Mean Ratings of Six Groups on 10 Evaluation Items	54
3.	Results of Univariate ANOVA for Six Groups	56
4.	Regrouping of Faculty According to Squared Distance From Centroids	57
5.	Mean Ratings for Tenured and Nontenured Faculty on 10 Evaluation Items	59
6.	Results of Univariate ANOVA for Tenured and Nontenured Faculty	60
7.	Mean Ratings for Professors, Associate Professors, and Assistant Professors on 10 Evaluation Items	61
8.	Results of Univariate ANOVA for Professors and Associate Professors	63
9.	Results of Univariate ANOVA for Associate Professors and Assistant Professors	64

Abstract of Dissertation Presented to the Graduate
Council of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

A COMPARISON OF THE TEACHING EFFECTIVENESS OF
TENURED AND NONTENURED FACULTY AS MEASURED
BY STUDENT EVALUATIONS

By

Lily M. Loretta James

June, 1975

Chairman: Kern Alexander

Major Department: Educational Administration and Supervision

The purpose of this study was to examine the null hypothesis that there were no differences in the ratings received by tenured and nontenured faculty in student evaluations of teaching effectiveness. This study further sought to determine if there were student evaluation measures which were closely associated with tenure status and if there were differences among the ratings of faculty members in the ranks of professor, associate professor, and assistant professor.

Study participants were 268 professors, associate professors, assistant professors, and instructors in the College of Arts and Sciences at the University of Florida, who had been evaluated by 7,901 students during the Winter Quarter of 1974.

Data were obtained through a study of publications and a study of records which were on file. The source of the

data was the instrument used by the College of Arts and Sciences in its student evaluations of faculty members.

The unit of analysis was the item mean for each faculty member. Teaching effectiveness was measured on the basis of the mean rating received by a faculty member on each of the 10 items on the student evaluation form.

A multivariate one-way analysis of variance was performed to test for differences among six groups identified by both their tenure status and professorial rank: tenured professors, tenured associate professors, tenured assistant professors, nontenured associate professors, nontenured assistant professors, and nontenured instructors. Simultaneous confidence intervals were established for the rating means. Data from all 10 items on the evaluation instrument were used to determine whether a discriminant analysis would make a distinction between tenured and nontenured faculty. Means were calculated for the ratings of tenured and nontenured faculty and for professors, associate professors, and assistant professors. Hotelling's T^2 test was used to test for differences between the ratings of (a) tenured and nontenured faculty, (b) professors and associate professors, and (c) associate professors and assistant professors.

The results of this investigation indicated that:

1. Student ratings of teaching effectiveness for tenured and nontenured faculty were significantly different at the .05 alpha level.

2. Student ratings of teaching effectiveness for professors and associate professors and for associate professors and assistant professors were different at the .05 level of significance.
3. Nontenured faculty members received better ratings than tenured faculty members; associate professors received better ratings than professors; and associate professors received slightly better ratings than assistant professors.
4. The discriminant analysis made a distinction between tenured and nontenured faculty by reclassifying as tenured all except one tenured faculty member and by reclassifying as nontenured all except one nontenured faculty member.
5. No definite patterns of difference on evaluation items by tenured and nontenured faculty were shown.

Based on the findings of this investigation, attention and further study should be given to four relevant concerns: (a) the teaching effectiveness of faculty members prior to and following the granting of tenure; (b) the granting of tenure for a specified period, with independent evaluations for tenure at the end of each set time period; (c) the weight attached to teaching effectiveness in tenure considerations; and (d) the role of student evaluation of teaching effectiveness in tenure considerations.

CHAPTER I

INTRODUCTION

The decade of the sixties saw tremendous concern over the use of student evaluations as a measure of teaching performance in universities and colleges throughout the United States. This concern was followed—during the late sixties and early seventies—by a reemergence of the questioning of the concept of tenure. Logical questions have historically been raised and can, therefore, be raised regarding the relationship between a faculty member's tenure status and his performance in his position, particularly as it may relate to teaching. It was to these kinds of concerns that this inquiry was addressed.

Largely because of the financial exigency which plagued many public institutions of higher education during the 1973-74 school year, greater attention is being focused upon tenure. Faculty members who have not yet acquired tenure are concerned about their loss of jobs and the inability to move up in the professorial ranks. Faculty members who do have tenure are appalled over the actual inexistence of job security and are demanding that tenure laws be revised. In a 1974 article concerning the dismissal,

en masse, of tenured faculty, Semas stated:

The University of Wisconsin sent layoff notices last May to 88 tenured faculty members on nine of its campuses, all of them formerly part of the Wisconsin State University system.¹

Traditionally tenure has provided a fairly unquestionable security for faculty members. It has shielded many persons from the whimsical designs of their superiors, has provided the protection that faculty members needed as they embarked upon unorthodox methods and championed unpopular ideas, and has guaranteed persons' having jobs from year to year.

Today, however, many people are questioning the value of tenure, particularly in view of the recent firings of large numbers of tenured faculty members. Some persons believe that tenure should be abolished, but others still support tenure wholeheartedly. Some persons believe that tenure laws are inadequate but that they should not be abolished because they do provide a measure of protection. In an unprecedented decision, on the other hand, the Virginia State Board for Community Colleges decided to no longer grant tenure at its 22 institutions. As reported in Intellect, this move was believed to be the first of its

¹Philip W. Semas, "Faculty Firings Soar as Slump Hits Enrollment," The Chronicle of Higher Education, 28 January 1974, p. 1.

kind in the country.² Thus, tenure has created a major debate among educators.

At least three of the issues surrounding tenure have developed since the beginning of 1973—a factor which gives evidence of the ever-growing complexity of tenure: (a) tenure as a part of the collective bargaining process, (b) the role of nontenured teachers in institutional government, and (c) the tenure component of women and minority group members. Other issues, which were practically dormant, have come alive with unparalleled impetus: (a) tenure as it relates to the allocation of institutional finances, (b) financial exigency and its concomitant job insecurity, (c) tenure quotas, (d) tenure and the demand for greater accountability, (e) due process in terminating nontenured faculty members, and (f) the allocation of decisional power between trustees and faculty in the termination process. But despite the problems and controversies surrounding it, the practice of granting tenure remains rather stable.

Although teaching effectiveness as determined by administrative superiors has been used for many years as a criterion for granting tenure at the University of Florida, teaching effectiveness as measured by students has only been given such consideration since 1972. The focus of this

²"Elimination of Faculty Tenure," Intellect 101 (March 1973):346.

study was, therefore, upon student evaluations and the tenure status of faculty members at the University of Florida.

The Problem

Statement of the Problem

The problem of this study was to examine the null hypothesis that there were no differences in the ratings received by tenured and nontenured faculty in student evaluations conducted in the College of Arts and Sciences at the University of Florida. In this examination attention was given to the questions listed below:

1. Were there student evaluation measures that were closely associated with tenure status?
2. Were there significant differences among the ratings of faculty members in the academic ranks of professor, associate professor, and assistant professor?

Delimitations and Limitations

In conducting the research the investigator observed the following delimitations and limitations:

1. The study was confined to the College of Arts and Sciences at the University of Florida.
2. Only those faculty members who were evaluated during the Winter Quarter of 1974 were used in the study.
3. Data were obtained through a study of records and publications.

Justification for the Study

One of the avowed purposes of tenure is to guarantee the student of having a competent, effective teacher; but there is no empirical evidence to support the general belief that tenured faculty members are more effective teachers than nontenured faculty members. Because of the lack of research on tenure and teaching effectiveness as measured by student evaluations, this study will serve as an addition to the body of existing knowledge on tenure and as a reference for administrators and faculty in tenure considerations.

Assumptions

The basic assumptions in this investigation were that the results obtained through student evaluations were accurate and that the instrument used to measure teaching effectiveness was valid and reliable.

Definition of Terms

Academic rank.—The University's official designation of a faculty member as professor, associate professor, assistant professor, or instructor.

Evaluation.—The consideration of evidence regarding teaching in the light of the value standards, the particular situation, and the goals of the student.

Evaluation instrument.—The instrument used by the

College of Arts and Sciences in its student evaluation of faculty members.

Faculty.—Professors, associate professors, assistant professors, and instructors in the departments of the College of Arts and Sciences.

Professorial rank.—A faculty member's classification as professor, associate professor, assistant professor, or instructor.

Teaching effectiveness.—The rating which students gave a faculty member on the evaluation instrument.

Tenure.—Permanent faculty membership.

Tenured faculty.—Faculty members who had been granted tenure prior to the Winter Quarter of 1974.

Tenure status.—The classification of a faculty member as tenured or nontenured.

Procedures

Study Participants

The subjects who were selected for this study were the 268 University of Florida College of Arts and Sciences professors, associate professors, assistant professors, and instructors who were evaluated by 7,901 students during the Winter Quarter of 1974.

Sources of Data

The sources of data were the instrument which was used by the College of Arts and Sciences in its student evaluation of faculty members and records which were on file in the Office of Academic Affairs. The instrument used for student evaluations in the College of Arts and Sciences is presented below:

EVALUATION FORM

QUESTIONS:

1. The instructor was enthusiastic when presenting course material.
2. The instructor seemed to be concerned with whether the students learned the material.
3. You felt that this course challenged you intellectually.
4. The instructor appeared receptive to new ideas and others' viewpoints.
5. The instructor generally stimulated class discussion.
6. The instructor attempted to cover too much material.
7. The homework assignments were too time consuming relative to their contribution to your understanding of the course material.
8. The instructor's presentations made for easy note-taking.
9. The direction of the course was adequately outlined.
10. You generally enjoyed going to class.

RATINGS:

1. if you strongly agree with the statement
2. if you agree with the statement
3. if you neither agree nor disagree with the statement
4. if you disagree with the statement
5. if you strongly disagree with the statement

Data Collection

Data were collected through the steps listed below:

1. The researcher obtained from the registrar's office a copy of the "Course and Teacher Evaluation" for the Winter Quarter of 1974, which was prepared and published jointly by the University of Florida Senate and Omicron Delta Kappa.
2. From records on file in the Office of Academic Affairs, the researcher ascertained the tenure status and academic rank of each faculty member included in the Winter, 1974, student evaluation of faculty members in the College of Arts and Sciences.

Data Treatment

1. Ratings were averaged for each faculty member who had been evaluated by more than one class.
2. A multivariate test of hypothesis was performed to determine whether there were differences in the mean vectors for groups that were identified by both their tenure status and professorial rank.
3. Because it was concluded that the mean vectors were different, simultaneous confidence intervals were constructed for the rating means of the groups.
4. Data from the 10 items on the evaluation instrument were used to find out whether a discriminant analysis would make a distinction between tenured and nontenured faculty.

5. The mean was calculated for the ratings of tenured and nontenured faculty and for professors, associate professors, and assistant professors on each of the 10 evaluation items.
6. Hotelling's T^2 test was used to test for differences between the ratings for tenured and nontenured faculty members; professors and associate professors; and associate professors and assistant professors.

Organization of the Study

This thesis is presented in six chapters. Chapter I presents the problem of the investigation in perspective. Chapter II gives a chronological discussion of the practice of granting tenure and outlines the requirements for receiving tenure at the University of Florida. Chapter III is devoted to a review of related literature and research. The data are presented in Chapter IV. Chapter V contains the analysis and discussion of the data. And Chapter VI presents the findings and conclusions of the investigation and recommendations for further study.

CHAPTER II

BACKGROUND OF THE PROBLEM

The revival of interest in tenure during the 1970s implied that the concept of tenure was one of a short duration. But the current surge of interest in tenure is not an indicator of its span of existence; for tenure dates back to the Middle Ages, when it served as a fortification against external dangers. In his description of tenure during the Middle Ages, Metzger stated:

Tenure in magisterial universities was not an attribute of occupancy of office, with set emoluments and prescribed functions, but an attribute of admission to a corpus, possessed of a legal personality and (in part through delegation, in part through usurpation) of considerable governmental power, of a kind of minisovereignty. Admission to this body was accomplished not through the contract of employment but through the crossing of certain qualificatory thresholds—the earning of degrees, the exhibition of certain prowesses, the acquisition of a license—over which the masters took command.¹

To continue as a member of the corpus one was not required to perform specific duties (an individual could remain a master in good standing even if he did not teach); instead, he only had to adhere to collegial comities. And one could

¹Walter P. Metzger, "Academic Tenure in America: A Historical Essay," in Faculty Tenure, ed. W. R. Keast (San Francisco: Jossey-Bass, 1973), p. 101.

11

not be expelled except by members of the body itself.² Metzger regarded this period as "Tenure as Privilege: The Era of the Master."³ He further delineated two succeeding periods, which he referred to as "Tenure as Time: The Age of the Employee" and "Tenure as Judiciality: The Age of the Professional."⁴

Tenure as it is known in the mid-1970s is an outgrowth of the second period: The Age of the Employee, with Harvard College in the fore and William and Mary, Yale, and Princeton following. It should be noted, however, that even during this period tenure was in an evolutionary process for two centuries. Among the many changes which occurred at this time, two are of utmost significance—as forerunners of the present system of tenure. The first change established a limit for the number of years that one could remain in a professorial rank:

'In 1860, it [the corporation] limited not only the time of each tutorial appointment but the time each teacher could spend within that rank, the maximum number of years being set at eight.'⁵

A second major transpiration was the Murdock case, which challenged the criteria for termination. Although the Supreme

²Metzger, "Academic Tenure," p. 101.

³Ibid., p. 94.

⁴Ibid., pp. 111, 135.

⁵Ibid., p. 119.

Judicial Court of Massachusetts deferred to the findings of the governing board under the main charge of neglect of duty, in Murdock, the Court held that three other charges did not justify dismissal of the professor.⁶ Thus, the way was paved for future challenges of the substantive criteria for termination.

The third period, The Age of the Professional, was signaled by the formation of the American Association of University Professors (AAUP). In 1913, 18 professors at Johns Hopkins University signed a letter and sent it to their colleagues at nine other major institutions, urging them to join in forming a national professors' association. Six hundred select persons accepted the invitation to become charter members, and in January, 1913, the AAUP was formed.⁷

Quoting from the Lovejoy Papers (Lovejoy was the framer of the Johns Hopkins letter), Metzger stated:

The Johns Hopkins call proposed that the new professional body undertake 'the gradual formulation of principles respecting the tenure of professional office and the legitimate ground for the dismissal of professors' and that it establish 'a representative judicial committee to investigate and report in cases in which freedom is alleged to have been interfered with by administrative authorities of any university or in which serious

⁶James Murdock, Appellant from a Decree of the Visitors of the Theological Institution of Phillips Academy, in Andover, 7 Pick. 303 (1828).

⁷Metzger, "Academic Tenure," p. 135.

and unwarranted injury to the professional standing and opportunities of any professor is declared to have occurred.⁸

For a long time professors had been concerned about their security through tenure and—especially since the controversy over evolution, in the 1870s—they had sought academic freedom. But what was now so astonishingly significant was the uniting of these two concerns in one professional plan of action.⁹

General Guidelines for Tenure-Granting Institutions

The basic criteria for granting tenure in postsecondary institutions were established by the American Association of University Professors and the Association of American Colleges (AAC) in their "1940 Statement of Principles" (which supersedes all of the preceding Statements) and have been endorsed by all of the major higher education organizations in the United States.

The general stipulations of the 1940 Statement are that (a) institutions provide tenure, (b) the exact terms under which a person is being hired be stated in writing, (c) the probationary period for teachers not exceed seven years, (d) all teachers have academic freedom, (e) in cases of factual disputes all teachers be granted hearings, and (f) that

⁸Metzger, "Academic Tenure," p. 136.

⁹Ibid.

the termination of tenured faculty members because of financial exigency be bona fide. The criteria, as presented by the AAUP and AAC, do not prescribe precise rules to be followed; hence, each institution, while adhering to the 1940 Statement, has a rather broad latitudinal range for operation.

Criteria for Determining Tenure Status
at the University of Florida

Subscribing to the general evaluative criteria established in the "1940 Statement of Principles," the University of Florida has provided in its Constitution specific criteria for the granting of tenure, for example:

Eligibility for tenure for faculty members in the ranks of professor, associate professor, and assistant professor, or in equivalent ranks, shall begin after five years of service in a tenure-earning position ¹⁰

As stated in the Constitution, nomination of a faculty member for tenure is an indication that one is and will be an asset to the University:

Nomination of a faculty member for tenure shall signify that a high degree of competence has been demonstrated and continuing employment of the faculty member will serve the best interests of the institution and the University System.¹¹

¹⁰"Constitution of the University of Florida" (as amended through March, 1972), art. 5, sec. 5.

¹¹Ibid.

The tenure-granting process at the University of Florida requires that a faculty member be (a) recommended for tenure by the appropriate department chairman or unit officer, (b) nominated by the President, (c) reviewed by the Chancellor of the State University System, and (d) approved by the Board of Regents. Receiving tenure at the University of Florida is a major achievement for a faculty member, for it means that he has earned permanent faculty membership and will be employed continuously by the University until, as stated in the Constitution, he (a) voluntarily leaves the institution or retires, (b) reaches the mandatory retirement age, (c) is dismissed for justifiable cause (as described in the Constitution), (d) is no longer needed because his position is discontinued or the activities in which he is engaged are terminated, or (e) he dies.¹²

While adhering to the general criteria, as stipulated by the AAUP and AAC, the University of Florida has, nevertheless, exercised its prerogative to utilize student evaluation as an aid in determining the tenure status of its faculty. Article V, Section 5, of the "Constitution of the University of Florida" provides that:

At the time a faculty member becomes eligible for tenure, the appropriate department chairman or of-

¹²"Constitution," art. 5, sec. 5.

ficer in charge of the unit concerned, after consultation with the tenured members of the department or unit, and after taking into account other considerations such as student evaluations in the case of instructional positions, shall nominate him for that status . . . or shall postpone such nomination and, in either case, shall inform him in writing.¹³

Not only is the use of student evaluation in the tenure-determining process mentioned in publications designed expressly for faculty members, but also in publications prepared primarily for students. In the 1972-73 Course and Teacher Evaluation Handbook, students were urged to take care in completing future teacher evaluation forms for two reasons. One of the reasons was "the fact that beginning this year [1972-73] the evaluations are being used by administrators as an 'input in decisions affecting salary, tenure, and promotions.'"¹⁴ A similar statement concerning the use of student evaluations was included in the 1973-74 Course and Teacher Evaluation Handbook: the "information is used by the administration to affect salary, promotion, and tenure" ¹⁵

Initiated at the height of the nationwide tenure controversy, the use of student evaluations in tenure con-

¹³"Constitution," art. 5, sec. 5.

¹⁴Trevette Wallace, ed., Course and Teacher Evaluation Handbook (Gainesville, 1973), p. 2.

¹⁵Mike Rollo, ed., Course and Teacher Evaluation Handbook (Gainesville, 1974), p. 2.

siderations at the University of Florida may be viewed as a response to the problems created by tenure. For the controversies surrounding tenure have generated institutional and organizational concern and study throughout the United States. Indeed, in a 1973 study on tenure, Freeman and Rossmeier reported:

In recent months almost every periodical on higher education has devoted at least one article to the subject of tenure. The New York Times, the Wall Street Journal, and countless lesser known newspapers are also carrying articles about the debate. There is also the proliferation of such books as The Tenure Debate, Tenure, Aspects of Job Security on the Changing Campus, and Tenure in American Higher Education On the national scale the Commission on Academic Tenure in Higher Education, sponsored by the AAUP and the Association of American Colleges, has just concluded a two-year study.¹⁶

But despite the proliferative writing and research on tenure, no all-encompassing solutions to the issues surrounding tenure have been discovered. Even the two-year study by the Commission on Academic Tenure has not provided satisfactory answers. In a survey of 200 nontenured assistant professors, Budig and Decker reported that the respondents found little solace in The Tenure Debate or in the report of the Commission on Academic Tenure.¹⁷

¹⁶Thomas M. Freeman and Joseph G. Rossmeier, A New Look at Tenure: A Management Imperative (East Lansing, Mich.: Michigan State University, 1973), p. 1, ERIC Document Reproduction Service No. ED 080 097.

¹⁷Gene A. Budig and C. Richard Decker, "Assistant Professors, Disillusioned by Tenure, Believe Unionization Offers Better Protection," Phi Delta Kappan 55 (October 1973):143-44.

Thus, with no precedents set or no generalizable conclusions formed, it has become necessary for each institution to search for its own solutions to problems created by tenure. It is in this vein that this comparison of the student ratings of tenured and nontenured faculty at the University of Florida is made.

CHAPTER III

REVIEW OF RELATED LITERATURE AND RESEARCH

This review is devoted to literature and research on tenure, teaching effectiveness, and student evaluations—with emphasis upon the interrelationships between and among these areas. The three major subdivisions of the review are entitled "Tenure," "Teaching Effectiveness," and "Student Evaluations."

Tenure

The involvement of students in the teacher evaluation process is a response to tenure controversies, which have been intensified by financial exigencies and the demand for greater accountability. According to Tolor, efforts to appraise teacher competence have assumed increasing urgency because of three major developments: (a) "the growing acceptance of the principle of accountability," (b) "the changing economic picture in which a combination of recessionary trends and inflationary pressures have produced a diminished tolerance on the part of most people for assuming heavy tax burdens," and (c) "the mounting concern with the issue of the relevance of the educational enterprise in a rapidly

changing society where traditional values and goals are constantly challenged."¹

A crucial factor concerning institutional financial need, however, is the stipulation in the "1940 Statement of Principles" that tenured faculty be released only in cases of bona fide financial exigency. Thus, whenever an institution is faced with financial problems, nontenured faculty members are the first to be released. And without evidence to support such action, no conclusions can be drawn about the teaching effectiveness of tenured versus nontenured faculty—that is, it may be true that the stipulation above protects competent faculty members. On the other hand, one might find that the stipulation above shelters incompetent faculty members at the expense of more competent ones.

The results of a study conducted by Bauer showed that faculty members viewed the tenure system as one that could lead to the protection of incompetence.² The purpose of Bauer's study was to compare the tenure process and faculty

¹Alexander Tolor, "Evaluation of Perceived Teacher Effectiveness," Journal of Educational Psychology 64 (February 1973):98.

²Roger D. Bauer, "Security Versus Freedom: Some Faculty Views on the Tenure Process," Phi Delta Kappan 54 (December 1972):280.

attitudes toward it at one campus of the University of California and one campus of the California State University System. Twenty faculty members from each campus were interviewed the year immediately after they received tenure. The key question posed to the interviewees concerned what each person felt was the most apparent weakness of the tenure appointment system at that time. The unanimous response was "the tenure system can be subject to abuse rather easily and it can lead to undesirable protection of incompetence or foster mediocrity."³

Discussing the tenure problem in a similar vein, Hughes stated, "What began as a system to protect professional civil liberties has become instead a reward for the few and a curse to the many."⁴ Further, in relating the United States' economic crisis to tenure, Hughes gave the following description:

Swollen populations in some areas with too few instructors, shrunken populations in other areas with a surplus of instructors—and an economy and a tenure system that preclude hiring and firing where needed.⁵

³Bauer, "Security Versus Freedom," pp. 279-80.

⁴Richard E. Hughes, "How to Solve the Tenure Problem: Departmental or College Professorship," College Management 9 (January 1974):7.

⁵Ibid.

The Tenure Debate

Indeed, the nation's economic plight—with its concomitant effects upon higher education—has led to a major debate among the proponents and the opponents of tenure. In examining the issues of the tenure debate, Shulman pointed out that "critics of the tenure system do not dispute the need for academic freedom, but they contend that the two can be separated and that academic freedom can be protected by other means."⁶

In addition, Mann reported that the results of a survey of American Association for Higher Education (AAHE) members showed that 64% of 1560 respondents felt that the tenure system needed major modifications, while only 15% felt that the tenure system was basically sound and 21% felt that it should be completely abandoned.⁷ Another survey, by Budig and Decker, showed that 200 nontenured assistant professors at 10 major public universities (mainly Big Ten schools) felt that tenure in no way protected their academic freedom. But 193 of them believed that tenure did protect the freedom of tenured professors. Two-thirds of the respondents, feeling more threatened from within academia than without, expressed concern about being at the mercy of their tenured

⁶Carol H. Shulman, "The Tenure Debate," in Research Currents (Washington, D. C.: ERIC Clearing House on Higher Education, 1971), p. 2, ERIC Document Reproduction Service No. ED 054 391.

⁷William R. Mann, "Is the Tenure Controversy a Red Herring?" Journal of Higher Education 44 (February 1973):86.

colleagues—that is, that tenured faculty members would determine their tenure fate.⁸

Further attestation to the tenure debate is the change in tenure policies from 1972 to 1974. A typical example is the change that has been effected in regard to the maximum length of the probationary period for tenure, as indicated through surveys. Surveys conducted by El-Khawas and Furniss showed that in 1972, among public and private 4- and 2-year colleges, only 0.4% of private 4-year colleges had as a maximum-length probationary period for tenure "9 or more years." In 1974, however, 1.7% of private colleges and 7.2% of private universities had a maximum probationary period of "9 or more years." Another significant finding was that in response to the question of whether full-time faculty could be granted tenure at an institution the percentages of the colleges which responded "yes" decreased from 1972 to 1974: in private 4-year colleges the decrease was from 94.1% to 93.2%; in public 2-year colleges from 68.3% to 66.9%; and in private 2-year colleges from 68.2% to 65.7%.⁹

⁸Budig and Decker, "Assistant Professors Believe Unionization Offers Better Protection," pp. 143-44.

⁹Elaine H. El-Khawas and W. Todd Furniss, Faculty Tenure and Contract Systems: 1972 and 1974 (Washington, D. C.: American Council on Education).

Tenure and Accountability

Despite the fact that public institutions are limited by law in what they can do to modify tenure systems, the public is demanding more returns for dollars invested in education. This demand for accountability has created greater interest in tenure. "Since 1967," Lieberman stated, "13 states have enacted legislation dealing with personnel evaluation, and such legislation has been introduced but not enacted in at least seven other states."¹⁰

Yet, while citizens are asking for evidence showing that they are getting what they are paying for, faculty members are holding on tightly to the jobs that they have. A few years ago—during more prosperous times—teachers were not overly concerned about dismissals, for they knew that teaching positions were available elsewhere. But the scarcity of jobs during the mid-1970s, the declining enrollments, rising costs, emphasis upon affirmative action, and tenure quotas have almost halted job mobility. It is these conditions that have created a dilemma for educational governing boards, for both the internal and external forces must be reckoned with.

Commenting on the soundness of public demand for accountability, Magarrell quoted Joanne Arnold as having

¹⁰Myron Lieberman, "Tenure: A New High-Priority Issue," Phi Delta Kappan 56 (March 1975):450.

said, "'The guy that's paying for it wants to know what he's getting—and I think that's good.'"¹¹ Arnold also said:

'In this age of consumerism and maybe a renewed zeal on the part of legislators and funders, at least the state-funded institutions are going to have to produce information showing what various programs cost, how resources are being used—not only money, but people'¹²

Teaching Effectiveness

For years educators have exhibited interest in teaching effectiveness. In 1944, for example, faculty members at the University of Washington were asked to complete a questionnaire on which one of the items was "Do you believe that administrative officers have adequate information concerning teaching efficiency?" Fourteen percent of the faculty responded in the affirmative, 31% expressed doubt, and 55% made their doubt emphatic. In the same questionnaire, the faculty was asked to list items that should govern promotion. The item mentioned most frequently was "teaching effectiveness." Yet when asked how teaching effectiveness could be measured, 18% of the faculty indicated that measurement would be difficult or impossible.¹³

¹¹Jack Magarrell, "Trying to Devise a Way to Measure the Benefits of Higher Education," The Chronicle of Higher Education, 1 April 1974, p. 12.

¹²Ibid.

¹³E. R. Guthrie, "The Evaluation of Teaching," Educational Record 30 (January 1949):110-11.

Research on teaching effectiveness dates back to the 1920s. McKeachie reported that the earliest research dealt with class size.¹⁴ This research was centered upon the question of whether small classes provided for more effective teaching than large classes. Among the first investigators were Edmonson and Mulder, who compared the performance of students matched for intelligence in a class of 109 with that of a class of 43 students.¹⁵ Edmonson and Mulder's research led to other investigations on class size. Subsequent research on teaching effectiveness examined other variables believed to be relevant to the teaching-learning process, for example, teaching experience, professorial rank, sex, age, and course grade.

But despite the emphasis placed upon research on teaching effectiveness and despite their concern about teaching effectiveness, educators have done little to promote effective teaching. Commenting upon this discrepancy between theory and practice, Wilson stated:

Everywhere, of course, there is lip service to the importance of good teaching; yet in few places is

¹⁴Wilbert J. McKeachie, "Research in Teaching: The Gap Between Theory and Practice," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), pp. 212-13.

¹⁵J. B. Edmonson and F. J. Mulder, "Size of Class as a Factor in University Instruction," Journal of Educational Research 9 (January 1924):1-12.

there systematic and rigorous attention to reinforcing, rather than obstructing, the values institutions claim to hold in this regard.¹⁶

Describing, in a similar manner, faculty attitudes toward teaching, Eble said:

Faculty members respect teaching and are somewhat interested in it, but comparatively few incline toward developing teaching as an art or [developing] themselves primarily as teachers The most accurate term I can find for the common faculty attitude is respect. It is not enthusiasm . . . not hostility.¹⁷

Wilson, in addition, indicated that both teachers and administrators were responsible for the low value placed upon teaching:

Both faculty and administration must accept [the] blame for the fact that consultantships, outside research grants and contracts, and publications score more points in the ratings of academicians than does dedicated and effective teaching.¹⁸

Teaching Effectiveness and Accountability

Universities and four-year colleges throughout the nation have customarily stressed the publish-or-perish

¹⁶Logan Wilson, "The Professor and His Roles," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), p. 104.

¹⁷Kenneth E. Eble, Professors as Teachers (San Francisco: Jossey-Bass, Inc., 1972), pp. 24-25.

¹⁸Wilson, "Professor and His Roles," p. 104.

syndrome in their hiring practices and tenure and salary considerations, often relegating teaching to a position that was secondary to research and publications. The results of a 1966 survey of all higher education institutions in the United States showed not only that teaching effectiveness was considered secondary in importance, but also that research and publications were being used as determinants of teaching ability. Astin and Lee reported:

It is clear from the data . . . that the professor's scholarly research and publication—not information based on classroom visits, systematic student ratings, student performance on examinations, and similar sources—are currently the primary considerations in evaluating his teaching ability.¹⁹

Indeed, for many years university and college administrators were free to place emphasis upon research and publications; the general public took no steps to interfere with their establishing of priorities. But in 1973 Whitfield and Brammer wrote, "The time is at hand when our higher institutions must demonstrate that quality instruction is their business or anticipate that their clients will go elsewhere."²⁰ Whitfield and Brammer pointed out that

¹⁹Alexander W. Astin and Calvin B. T. Lee, "Current Practices in the Evaluation and Training of College Teachers," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), p. 298.

²⁰Raymond P. Whitfield and Lawrence M. Brammer, "The Ills of College Teaching: Diagnosis and Prescription," Journal of Higher Education 44 (January 1973):1.

although colleges and universities had been jolted by a variety of criticism in recent decades, the students and the public—whose support was essential—had not yet delivered the blow to which institutions of higher education were most vulnerable.²¹ They explained:

The academic Achilles' Heel is a low level of teaching performance which the institutions have not seriously tried to improve. Neither hiding this neglect behind a historically nurtured professional mystique nor continued hope in the patience of the clientele will any longer serve as adequate defense.²²

In addition, as recently as 1974, Scully reported that the Group for Human Development in Higher Education was calling for "campus teaching institutes that would coordinate a wide variety of activities designed to give teaching an equal place with research and publication in higher education."²³

Lay citizens, however, have now demonstrated concern in regard to faculty members' teaching performance. When they ask, "What are we getting for our money?" they are not inquiring about the amount of research teachers have done or the number of books teachers have written; they are asking, instead, how successful faculty members are in their teaching.

²¹Whitfield and Brammer, "Ills of College Teaching," p. 1.

²²Ibid.

²³Malcolm G. Scully, "Improve Teaching, Prevent Stagnation, Group Urges," The Chronicle of Higher Education, 18 March 1974, p. 1.

And in this thrust for accountability the public is particularly aiming its criticism at tenure systems and tenured teachers, expressing the belief that tenure laws shelter incompetent teachers, who receive unmerited tax dollars.

Teaching Effectiveness and Tenure

Although it is generally believed that there is a significant causal relationship between teaching effectiveness and tenure, some educators have pointed out that such a relationship is more of an assumption than a reality. Whitfield and Brammer stated:

Despite arguments that promotion and tenure depend upon good teaching, conditions generally permit the professor to improve or neglect his teaching as he chooses. One inevitable consequence is the temptation for professors to attain a minimal level of teaching competence needed for tenure and change little thereafter.²⁴

Further, in matters concerning tenure, pay, and promotion some faculty members regard the teaching function and the research functions as equals. Costin et al. reported, "Some faculty members claim that since good teaching and good research go hand in hand, it is sufficient to reward the good researcher."²⁵

²⁴Whitfield and Brammer, "Ills of College Teaching," p. 2.

²⁵Frank Costin, William T. Greenough, and Robert J. Menges, "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," Review of Educational Research 41 (December 1971):511-35.

Teaching Effectiveness as Defined
by the Public in the Mid-1970s

But the public is not interested in the traditional methods of evaluating teaching effectiveness; on the contrary, lay citizens in the mid-1970s are defining teaching effectiveness as student achievement. As Lang has pointed out, whereas employees are usually evaluated by their quantitative and qualitative output, in schools the critical output is not the teacher's performance but student achievement. And "the public," he said, "is interested in student achievement," despite a widespread belief that measures to determine the relationship between teacher behavior and student achievement have not been adequately established.²⁶

Eble stated, however, that although college teachers may claim that the relationship between teacher behavior and student achievement has not been fully established, ample research has indicated what constitutes effective teaching:

What most college teachers are not aware of is the great number of careful scholarly investigations which have tried to find out what does constitute effective and ineffective teaching Were the results of these investigations at wide variance with one another, the skeptic might find support for his distrust of knowing anything about the subject Whether one is looking for the general qualities of the effective teacher or for

²⁶Theodore H. Lang, "Teacher Tenure as a Management Problem," Phi Delta Kappan 56 (March 1975):460.

more specific characteristics of effective teaching, research studies provide reasonably consistent answers.²⁷

Public impatience may, then, be attributable in part to faculty members' continuous search for more indices of teaching effectiveness. The impatience may also be an outgrowth of the general uncertainty of the times—uncertainty about government, the economy, and foreign relations. On the other hand, the current public demand for evidence of teaching may be the result of a combination of other factors: increased taxation, spiraling educational costs, inflation, and unemployment. But whatever the reasons may be, it is clear, according to Lang, that "there is a general lack of confidence in the teaching profession."²⁸ And although the public is pressing its demand for accountability—13 states having enacted accountability legislation which mandates evaluation procedures for educational personnel—Lang also asserted that "the public is no longer willing to rely on professional educators to evaluate themselves."²⁹

²⁷Kenneth E. Eble, The Recognition and Evaluation of Teaching (Salt Lake City: Project to Improve College Teaching, 1970), p. 9.

²⁸Lang, "Teacher Tenure," p. 459.

²⁹Ibid.

Student Evaluation as a Measure of Teaching Effectiveness

An alternative or supplement to evaluations by administrators and colleagues is student evaluation. Although there are negative factors to be considered in the use of student evaluation, salient arguments have been presented in support of student evaluation as a measure of teaching effectiveness. Howe stated, for example:

Students do pay for the instruction they receive; they are not simply a necessary evil to be tolerated as part of the educational endeavor, but are the purpose of it.) The opinions of those who eat the pudding certainly ought to be considered if we wish to know how the pudding tastes.³⁰

In a similar manner, Johnson quoted a Yale student as having said:

'It is often the student who knows best when he cannot understand or already knows what is being discussed.

It is the student who knows when a course is stimulating him to learn more about a subject or whether it is boring him to death.

It is the student who can best formulate those fundamental and personal questions so bothering him that he cannot readily proceed to other academic matters.

It is the student who can best evaluate when he is beginning to integrate the process of learning with the problems he continually confronts in life.'³¹

³⁰Harold Howe, "Less Teaching, More Conversation," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), p. 260.

³¹James A. Johnson, "Instruction: From the Consumer's View," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), pp. 289-90.

If educators really want to evaluate teaching effectiveness, if they really want to find the answers; "they must," Johnson contended, "be willing to involve students and encourage honest answers."³²

Student Evaluations

One of the most sensitive areas in higher education is that of evaluation of instruction. In most colleges and universities, teaching has been and still is being judged on the basis of research and publication, hearsay, subjective peer ratings, self-evaluations, student opinions, systematic student ratings, or a combination of two or more of these bases. The increased emphasis on accountability, the pressures of tenure systems, the surplus of college teachers, limited economic resources, and a rapidly changing society; however; have collectively created a demand for more widespread use of systematic student evaluations in measuring teaching effectiveness.

According to Gustad, Gage identified three significant reasons for evaluating teaching effectiveness: (a) "to provide a basis on which a number of administrative decisions can be made," (b) "to provide a basis for self-improvement by the faculty," and (c) "to provide a criterion that can

³²Johnson, "Instruction," p. 290.

be employed in research on teaching and learning."³³ It was McKeachie, however, who expressed the idea that students—the "consumers" of courses—might provide the best judgment of the criteria for evaluating instruction.³⁴ Viewing the emphasis upon student evaluation of faculty as an outgrowth of the student unrest of the 1960s, Eble pointed out that "evaluation is one of the few formal ways students have of indicating to the faculty what they want."³⁵

Use of Student Evaluations

Regardless of its novel appearance, the concept of student evaluation of faculty members is not new. As Kent indicated:

The spate of 'report cards' from students to teachers during the mid-1960's was attended by a considerable amount of fanfare in the newspapers: 'A NEW STUDENT REVOLT'; 'FALLOUT FROM BERKELEY'; 'THE BOOKWORM TURNS'; 'SHAPE UP, PROF!' ³⁶

³³John W. Gustad, "Evaluation of Teaching Performance: Issues and Possibilities," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), pp. 266-67.

³⁴Wilbert J. McKeachie, Teaching Tips: A Guide for the Beginning Teacher, 6th ed. (Lexington, Mass.: D. C. Heath and Co., 1969), p. 213.

³⁵Eble, Professors as Teachers, p. 73.

³⁶Laura Kent, "Student Evaluation of Teaching," in Improving College Teaching, ed. C. B. T. Lee (Washington, D. C.: American Council on Education, 1967), p. 313.

Yet, Kent said that the idea of student evaluation of teaching was not first introduced during the 1960s:

In spite of the bemusement of the headline writers, the notion of having students rate the effectiveness of their instructors is not . . . new. The University of Washington has conducted campuswide evaluations since 1925. The University of Texas and Purdue were experimenting with such evaluations, though on a more limited scale, at about the same period.³⁷

In addition, Kent stated that as early as 1924 Harvard had initiated the publishing of a course and teacher evaluation booklet: A Confidential Guide to Courses.³⁸ What Kent said was new, however, were the manner in which and the purpose for which the ratings were being used in the sixties:

What is so 'revolutionary' and newsworthy about the student rating programs being initiated or proposed today is that the results are no longer the private affairs of the instructor Now the teacher's classroom behavior is being subjected to the scrutiny of faculty committees on promotion and tenure, and student ratings are being used for 'administrative purposes.'³⁹

Expressing her views in a prediction concerning the use of student evaluations, Samalonis concluded:

Student ratings will be an additional factor in the evaluation of college teachers, and . . . these ratings will help to reward the dedicated

³⁷Kent, "Student Evaluation," p. 313.

³⁸Ibid., p. 328.

³⁹Ibid., p. 314.

college teacher for his outstanding contribution to the students and to the college.⁴⁰

Although Kent stated clearly that there was nothing new about the use of student evaluation in measuring teaching effectiveness, she also indicated that student evaluation of teaching effectiveness was not a widespread practice in institutions of higher education in the United States. In fact, she reported that in 1966 student ratings were being used to a lesser extent than they had been used in 1961.⁴¹ In her 1966 study Kent found that 47.6% of 1,110 responding higher education institutions did not use student ratings at all. And among the institutions that did use student ratings of teaching effectiveness, university agriculture and business colleges used them most extensively, and junior colleges ranked next in frequency of use.⁴²

But in the results of a more recent study, Bauer stated, "Student evaluations are widely used to assess teaching effectiveness."⁴³ However, he further indicated

⁴⁰Bernice Samaloni, "Ratings by Students," Improving College and University Teaching 15 (Winter 1967):11.

⁴¹Kent, "Student Evaluations," p. 317.

⁴²Ibid.

⁴³Bauer, "Security Versus Freedom," p. 280.

that the analysis of his data showed that there was "great variance of opinion regarding use, selection, evaluation, and application of the student evaluation."⁴⁴

Faculty Views Regarding Student Evaluations

Generally, faculty members view the evaluation of teaching effectiveness as essential in making decisions concerning promotion, raises, and tenure. Teachers' favorable attitudes toward evaluation of teaching effectiveness were indicated in the results of a survey conducted by Blai: virtually all of the 1,105 faculty members—a sample from seven colleges and universities—stated that teaching effectiveness should be a major factor in pay and promotion matters.⁴⁵ But faculty attitudes appear to be positive only in regard to evaluation of teaching as performed by their administrative superiors. Their views about student evaluation in job advancement and tenure decisions are, on the whole, negative. Commenting on faculty members' objection to student evaluations, Bauer stated, "A major objection revolves around the qualification of students to

⁴⁴Bauer, "Security Versus Freedom," p. 280.

⁴⁵B. Blai, "Faculty-Ranked Importance of Several Criteria in Pay-Promotion Matters," Scientia Paedagogica Experimentalis 9 (1972):5-17.

judge the effectiveness of a given professor in terms of the university's overall objectives."⁴⁶

Another reason for faculty objection to the use of student evaluations concerns the construction of the evaluation instrument. The opponents of student evaluation have said that students lack the qualifications to construct the instrument. Regarding this claim, Costin et al. said, "Some faculty members will frequently challenge the administration and potential use of student ratings of instruction no matter who prepares the forms."⁴⁷ Further, Costin et al. stated that the opponents of student evaluations claim that (a) "ratings will favor an entertainer over the instructor who gets his material across effectively," (b) "ratings are highly correlated with expected grades," and (c) "students are not competent judges of instruction since long-term benefits of a course may not be clear at the time it is rated."⁴⁸ But, Costin et al. pointed out, the proponents of student ratings have contended that the validity of the objections above is testable. In addition, faculty members

⁴⁶Bauer, "Security Versus Freedom," p. 280.

⁴⁷Costin, Greenough, and Menges, "Student Ratings of College Teaching," p. 511.

⁴⁸Ibid.

who favor student evaluations have stressed the value of these ratings to both the individual teacher and the entire college.⁴⁹

Students' Concepts of the Effective Teacher

Students, the "consumers" of education, have always expressed their opinions about effective and ineffective teachers; these opinions, however, have often been exchanged among friends in informal settings and casual conversations. And in this manner a valuable source of feedback to teachers and administrators has not been fully utilized. Nevertheless, some formal procedures have been used to determine those characteristics that students have considered typical of the effective teacher.

Crawford and Bradshaw asked students to describe the most effective college teacher they had ever known and to write a theme describing the characteristics most essential to his being an effective teacher. Of the 13 descriptive statements sorted out, the four most frequently mentioned characteristics were (a) thorough knowledge of subject, (b) well-planned and well-organized lectures, (c) lively, enthusiastic interest in teaching, and (d) friendly

⁴⁹Costin, Greenough, and Menges, "Student Ratings of College Teaching," p. 511.

and willing to help students.⁵⁰

In a similar study Musella and Rusch obtained responses consistent with those of Crawford and Bradshaw. Seniors at the State University of Albany were asked to describe the teacher behaviors that had been most influential in promoting their thinking. The characteristics mentioned most frequently by the 394 respondents were (a) enthusiastic attitude toward the subject, (b) sympathetic attitude toward students, (c) effective use of questions, (d) skill in presenting material, and (e) knowledge of subject.⁵¹

A frequently heard argument is that students—because of their youth, inexperience, and short-term perspectives—are incapable of establishing adequate criteria for evaluating teacher effectiveness. Drucker and Remmers found evidence that seemingly refuted this charge. Using Purdue alumni who had been out of college for 10 years and undergraduates who were currently enrolled at Purdue, they found substantially positive relationships between the

⁵⁰P. L. Crawford and H. L. Bradshaw, "Perception of Characteristics of Effective University Teachers: A Scaling Analysis," Educational and Psychological Measurement 28 (Winter 1968):1080-81.

⁵¹D. Musella and R. Rusch, "Student Opinion on College Teaching," Improving College and University Teaching 16 (Spring 1968):137-38.

ratings of teachers by both groups. Correlation coefficients for 10 traits ranged from .40 to .68.⁵²

Reliability of Student Evaluations

Student evaluations, said Wood et al., have been criticized on the grounds that they represent only the student's perspective on teacher performance. Yet, they stated, "the empirical studies of student ratings suggest that they have a high degree of reliability."⁵³ In one of the earlier studies on reliability, Guthrie found that students' judgments of their teachers were quite stable from year to year. Comparing students' 1949 rankings of faculty members with their 1950 rankings, Guthrie found a correlation of .87. When he compared students' 1951 rankings of teachers with those of 1949, he found a correlation of .89.⁵⁴ A significant factor about the two sets of comparisons above is that in making the second set of comparisons, 1949 and 1951, Guthrie eliminated a major threat to the external validity of his study: selection and treatment interaction. The

⁵²A. J. Drucker and H. H. Remmers, "Do Alumni and Students Differ in Their Attitudes Toward Instructors?" Journal of Educational Psychology 42 (March 1951):138-42.

⁵³Kenneth Wood, Arnold S. Linsky, and Murray A. Straus, "Class Size and Student Evaluations of Faculty," Journal of Higher Education 45 (October 1974):526.

⁵⁴E. R. Guthrie, The Evaluation of Teaching: A Progress Report. (Seattle: University of Washington, 1954), p. 4.

teachers who were rated in 1950 requested student evaluation; in 1951 faculty members were not given the opportunity to decide whether they would be evaluated; all were included.

Using a forced-choice rating to obtain students' evaluations of their teachers, Lovell and Haner found that the correlation between ratings made two weeks apart was .89. Split-half reliability was $r = .88$ (corrected).⁵⁵ In a more recent study, Costin found r s ranging from .70 to .87 between midsemester and end-of-semester ratings of teaching assistants for four dimensions measured: skill, structure, feedback, and rapport. The r for the fifth dimension—group interaction—was .48.⁵⁶

In regard to the internal consistency of student ratings, Maslow and Zimmerman reported r s from .77 to .94 when the ratings of students in a given class were randomly paired.⁵⁷ Comparing the mean odd-item ratings on their forced-choice instrument with the mean even-item

⁵⁵G. D. Lovell and C. F. Haner, "Forced-Choice Applied to College Faculty Rating," Educational and Psychological Measurement 15 (Autumn 1955):303.

⁵⁶Frank Costin, "A Graduate Course in the Teaching of Psychology: Description and Evaluation," Journal of Teacher Education 19 (Winter 1968):430-31.

⁵⁷A. H. Maslow and W. Zimmerman, "College Teaching Ability, Scholarly Activity and Personality," Journal of Educational Psychology 47 (March 1956):187.

ratings, Lovell and Haner found that the r was .79. Corrected by the Spearman-Brown prophecy formula, the r increased to .88.⁵⁸ The results of the studies discussed above do support a high degree of reliability for student evaluations.

Validity of Student Evaluations

Research on the validity of student evaluations does not indicate such positive or concrete results as that on reliability. McKeachie proposed his idea of having students, the "consumers" of courses, provide the criteria for evaluating teacher performance in view of his realization that "the validity of student ratings has been the chief bone of contention between those for and those against the use of student ratings."⁵⁹ McKeachie hypothesized that the validity of student ratings could be determined on the basis of the extent to which students' subjective criteria matched faculty members' teaching goals.⁶⁰ In a comparative study of student and teacher evaluations, Centra found, however, only a modest relationship:

A comparison of students' ratings of instruction with teachers' self-reported ratings in over 300

⁵⁸Lovell and Haner, "Forced-Choice," p. 297.

⁵⁹McKeachie, Teaching Tips, p. 213.

⁶⁰Ibid.

classes at five colleges disclosed a modest relationship between the two sets of evaluations. In addition to the general lack of agreement between self- and student evaluations, there was also a tendency for teachers as a group to give themselves better ratings than their students did.⁶¹

In an earlier study—one whose results, they said, were less convincing than they had hoped for—McKeachie, Lin, and Mann attempted to determine some of the bases on which students made their judgments. They did find that teachers whose students rated them high on "rapport" ("warmth") were effective on measures of students' "critical thinking."⁶² McKeachie and Solomon had presented some evidence that students of highly rated instructors elect more courses in the same field. But the data analysis had not indicated that there was a very consistent relationship between student evaluations and student interest through the election of advanced courses.⁶³

Another method used in attempts to validate student ratings is the comparison of these ratings with colleagues' ratings of teachers. In 1949 and 1954 Guthrie found positive

⁶¹John A. Centra, "Self-Ratings of College Teachers: A Comparison with Student Ratings," Journal of Educational Measurement 10 (Winter 1973):293.

⁶²Wilbert J. McKeachie, Yi-Guang Lin, and William Mann, "Student Ratings of Teacher Effectiveness: Validity Studies," American Educational Research Journal 8 (May 1971):444.

⁶³Wilbert J. McKeachie and D. Solomon, "Student Ratings of Instructors: A Validity Study," Journal of Educational Research 51 (January 1958):380-81.

rs of .48 and .63 respectively between students' and colleagues' ratings of the same teachers.⁶⁴ Maslow and Zimmerman found a .69 correlation in their comparison of students' and colleagues' ratings; they also found that students tended to equate good teaching with "good personality," whereas colleagues equated good teaching with "creativity."⁶⁵ Although the studies on students' and colleagues' ratings have shown positive correlations, Costin et al. pointed out that the associations have not been significant enough to support the contention that student ratings, used alone, have a contribution to make to the evaluation of teaching.⁶⁶

Other investigations of the validity of student ratings have compared student evaluations with teachers' years of experience. In an early study Heilman and Armentrout reported that there was no significant relationship between teachers' experience and student ratings.⁶⁷ Guthrie also

⁶⁴E. R. Guthrie, "The Evaluation of Teaching," Educational Record 30 (April 1949):113; Guthrie, The Evaluation of Teaching, p. 5.

⁶⁵Maslow and Zimmerman, "College Teaching Ability," p. 187.

⁶⁶Costin, Greenough, and Menges, "Student Ratings of College Teaching," pp. 516-17.

⁶⁷J. D. Heilman and W. D. Armentrout, "The Rating of College Teachers on Ten Traits by Their Students," Journal of Educational Psychology 27 (March 1936):209, 215.

reported a negligible correlation of .13 between student ratings and faculty members' years of teaching experience.⁶⁸ Downie, on the other hand, found that full professors received higher ratings than those in other ranks—especially on such characteristics as effectiveness of presentation and sense of humor.⁶⁹ And Gage reported that full professors and associate professors received significantly higher ratings than assistant professors and instructors. In addition, he—just as Remmers and Elliott—found that teachers of lower-level courses consistently received less favorable average ratings than teachers of advanced courses.⁷⁰ However, Guthrie had reported earlier that underclassmen and graduate students tended to agree strongly in their evaluations of teaching effectiveness. He found a correlation of .73.⁷¹

⁶⁸Guthrie, The Evaluation of Teaching, pp. 7-8.

⁶⁹N. W. Downie, "Student Evaluation of Faculty," Journal of Higher Education 23 (December 1952):496.

⁷⁰N. L. Gage, "The Appraisal of College Teaching," Journal of Higher Education 32 (January 1961):17; H. H. Remmers and D. N. Elliott, "The Indiana College and University Staff-Evaluation Program," School and Society 70 (July-December 1949):170.

⁷¹Guthrie, The Evaluation of Teaching, p. 4.

A question which has often been posed concerns the relationship of student grades to student evaluations. The typical a priori belief is that the higher the grade the student receives or expects to receive, the better the teacher's rating will be. Although some investigators—including Russell and Bendig and Stewart and Malpass⁷²—found low correlations between students' grades and their evaluations of teachers, the consensus among researchers is that there is no relationship between students' grades and their ratings of teachers. This opinion is based on the results of several studies—among them studies conducted by Guthrie and by Voeks and French⁷³—which have indicated that there is no correlation between student achievement and student evaluations.

In addition, empirical investigations, for example,

⁷²H. E. Russell and A. W. Bendig, "Investigation of the Relations of Student Ratings of Psychology Instructors to Their Course Achievement When Academic Aptitude Is Controlled," Educational and Psychological Measurement 13 (1953):634; C. T. Stewart and L. F. Malpass, "Estimates of Achievement and Ratings of Instructors," Journal of Educational Research 59 (April 1966):348.

⁷³Guthrie, The Evaluation of Teaching, pp. 6-7; V. W. Voeks and G. M. French, "Are Student-Ratings of Teachers Affected by Grades?" Journal of Higher Education 31 (June 1960):330-32.

those of Edmonson and Mulder; Guthrie; and Rohrer,⁷⁴ have led educational researchers to the belief that there is no relation between class size and student ratings of teachers.

Finally, in regard to the role of sex as related to student ratings, the results of several studies—Heilman and Armentrout; Downie; and Lovell and Haner⁷⁵—have indicated that there are no significant differences in female and male students' ratings of teachers or in the student ratings received by male and female teachers.

As noted from the outset, studies on the validity of student evaluations have not provided concrete answers. Thus, many of the questions concerning validity remain rather debatable. And it is these questions that lead the investigator on—in search of new answers and in search of evidence to substantiate the findings of earlier researchers.

Inconclusive Evidence Regarding Student Ratings

Although much literature has been written and many studies undertaken on student evaluations, much of the data

⁷⁴Edmonson and Mulder, "Size of Class as a Factor," p. 10; Guthrie, The Evaluation of Teaching, p. 6; J. H. Rohrer, "Large and Small Sections in College Classes," Journal of Higher Education 28 (May 1957):279.

⁷⁵Heilman and Armentrout, "The Rating of College Teachers," pp. 215-16; Downie, "Student Evaluation of Faculty," p. 503; Lovell and Haner, "Forced-Choice," p. 301.

available and many of the results obtained are inconclusive. McKeachie stated, "Numerous articles have been published, arguing for or against the use of student ratings. However, the research is scattered."⁷⁶ In the same manner, Tolor asserted that "despite numerous articles dealing with teaching effectiveness . . . much of the literature is inconclusive or offers conflicting results."⁷⁷

Nevertheless, Tolor pointed out that the search for answers concerning the evaluation of teaching effectiveness has continued:

The attempt to assess the quality of teaching has a long history and is, of course, intimately associated with issues of value judgments and measurement problems, including the most vexing one of criteria adequacy. Yet the search for the 'good teacher' has continued unabated, both in the research arena and in every community where educational facilities are staffed.⁷⁸

Indeed, the search has continued, and it will continue; for education has no purpose except it be to lead men on in quest of truth. Lee so aptly stated:

Each of us comes to this inquiry [on improving college teaching] with a different background and with different biases and commitments. But all of us are engaged in this enterprise as a rigorous intellectual activity, for such an inquiry requires not only organized reflection but also empirical study ⁷⁹

⁷⁶McKeachie, Teaching Tips, p. 213.

⁷⁷Tolor, "Perceived Teacher Effectiveness," p. 98.

⁷⁸Ibid.

⁷⁹C. B. T. Lee, ed., Improving College Teaching (Washington, D. C.: American Council on Education, 1967), p. 1.

CHAPTER IV

PRESENTATION OF DATA

The 7,901 student ratings for 268 faculty members in the College of Arts and Sciences at the University of Florida were the data used in this study. Class evaluations were limited to undergraduate courses and to 500-level graduate courses, which undergraduates were allowed to enroll in after obtaining special permission.

Statistical Analysis

The unit of analysis in this investigation was the item mean for each teacher. Among the 268 subjects used in the study, 60 had been evaluated by more than one class. For these 60 faculty members, the weighted mean was used, combining the groups.

Initially eight groups—defined by both their tenure status and professorial rank—were allowed for. The eight groups in which faculty members could have appeared were: Group 1, tenured professors; Group 2, tenured associate professors; Group 3, tenured assistant professors; Group 4, nontenured associate professors; Group 5, nontenured professors; Group 6, nontenured assistant professors; Group 7, tenured instructors; and Group 8, nontenured instructors.

The groups and the number of persons in each group are shown in Table 1.

Table 1
Faculty Members Classified by Group

	1	2	3	4
T	Group 1 N = 59	Group 2 N = 66	Group 3 N = 14	Group 7 N = 0
NT	Group 5 N = 0	Group 4 N = 9	Group 6 N = 116	Group 8 N = 4

T = Tenured
NT = Nontenured

1 = Professor
2 = Associate Professor
3 = Assistant Professor
4 = Instructor

Although allowance was made for eight groups, only six cells had entries. There were no nontenured professors and no tenured instructors. The six groups which had entries were subjected to statistical analysis.

Multivariate Analysis of Variance

A multivariate one-way analysis of variance was used to test for differences among the six groups. A one-way design was used because tenure status and rank were treated as one variable, since an examination of the data had shown

that status and rank were interrelated. The use of a multivariate analysis of variance (MANOVA), instead of an analysis of variance (ANOVA), provided for the comparison of groups in a more powerful test. Through MANOVA the mean ratings on all 10 evaluation items were compared simultaneously; this simultaneous comparison considered the possible interaction of the 10 variables with each other. And in this manner ratings on the evaluation instrument as a whole were analyzed. Had the purpose of the investigation been to point out the face-value differences among the groups on single evaluation items, an analysis of variance would have been sufficient. But any conclusions based on such an analysis could not have been very meaningful because no rating item would have measured purely what it was supposed to measure and nothing else.

Through the utilization of the Statistical Analysis System (SAS) Package (using procedure regression), the results of the MANOVA showed that the six groups were different at the .05 level of significance. In the test for differences among the groups, the value of F was found to be $F(50, 1157) = 1.48534$, $p > 0.0169$. Because this difference was significant, attention was directed toward finding where the differences were.

Although the mean ratings for the six groups on 10 evaluation items (see Table 2) showed the differences in group ratings, these ratings alone could not be regarded

Table 2

Mean Ratings of Six Groups
on 10 Evaluation Items

Group	Rating									
	1	2	3	4	5	6	7	8	9	10
1	1.86	2.20	2.15	2.43	2.70	3.11	3.50	2.44	2.14	2.54
2	1.64	1.97	1.85	2.09	2.36	3.19	3.65	2.38	1.95	2.11
3	1.63	1.90	2.13	2.08	2.24	3.45	3.55	2.24	1.91	2.30
4	1.71	1.95	1.88	1.87	2.17	3.20	3.69	2.45	1.92	2.13
6	1.73	1.94	1.94	2.07	2.29	3.22	3.66	2.27	2.05	2.26
8	1.77	1.93	2.39	2.17	2.40	3.45	3.60	2.58	2.24	2.20

as true indicators of those groups that were significantly different from each other. As a guide in determining where to look for differences among the groups, the univariate analysis showed that ratings 2, 3, 4, 5, and 10 had significance levels below .05 (see Table 3). On the other five ratings—1, 6, 7, 8, and 9—the significance levels were greater than .05, indicating that group differences on these ratings were not significant.

To find out where the significant differences were among the groups, simultaneous 95 percent confidence intervals were set up for the means for ratings 2, 3, 4, 5, and 10. These intervals, however, failed to show where the differences were.

Discriminant Analysis

The data were then subjected to a discriminant analysis, by which each individual was reclassified as a member of one of the six groups on the basis of his similarity to the entries in that group. An individual's similarity to a group's members was determined by examining the distance of the individual's point from the centroid, mean point, of each group; each subject was placed into that group to whose centroid his point was closest. Table 4 shows the number of faculty members who were placed into each group through the discriminant analysis. The discriminant analysis regrouped each of the 268 subjects according to his ratings

Table 3
Results of Univariate ANOVA for Six Groups

<u>Rating No.</u>	<u>Sign. Level</u>
1	0.2414
2	0.0454*
3	0.0086*
4	0.0009*
5	0.0023*
6	0.5068
7	0.5546
8	0.5680
9	0.1465
10	0.0086*

*Significant at .05 alpha level

Table 4
 Regrouping of Faculty According to
 Squared Distance From Centroids

From Group	Into Group					
	1	2	3	4	6	8
1 N = 59	29	15	15	0	0	0
2 N = 66	11	43	11	1	0	0
3 N = 14	4	3	7	0	0	0
4 N = 9	0	1	0	2	3	3
6 N = 116	0	0	0	7	84	25
8 N = 4	0	0	0	0	1	3

on the 10 variables on the student evaluation instrument.

Summary Measures in the Treatment of Data

As one of the summary measures in the treatment of the data, the ratings for the three groups of tenured faculty members (Groups 1, 2, and 3) and the ratings for the three groups of nontenured faculty members (Groups 4, 6, and 8) were averaged (see Table 5). Hotelling's T^2 test was then used to determine the differences between tenured and nontenured faculty ratings.

In the test for differences in the ratings for tenured and nontenured faculty, the two groups' ratings were found to be different at the .05 level of significance, $F(10,257) = 2.11810$, $p > 0.0233$. The univariate ANOVA showed the evaluation items on which the mean ratings for the two groups were significantly different (see Table 6).

Another summary measure was used to test for rating differences in professorial ranks. The means were calculated for the ratings received by professors, associate professors, and assistant professors (see Table 7) on the 10 evaluation items. Hotelling's T^2 test was used to ascertain differences between the ratings of professors and associate professors and differences between the ratings of associate professors and assistant professors.

The ratings of professors and associate professors were found to be different at the .05 level of significance,

Table 5
Mean Ratings for Tenured and Nontenured Faculty
on 10 Evaluation Items

Rating No.	T	NT
	(Groups 1, 2, 3)	(Groups 4, 6, 8)
1	1.73	1.73
2	2.06	1.94
3	2.00	1.95
4	2.23	2.07
5	2.49	2.29
6	3.19	3.22
7	3.58	3.66
8	2.39	2.30
9	2.03	2.04
10	2.31	2.25
	Tenured N = 139	Nontenured N = 129

Table 6
Results of Univariate ANOVA for Tenured
and Nontenured Faculty

<u>Rating No.</u>	<u>Sign. Level</u>
1	0.5758
2	0.0178*
3	0.3152
4	0.0037*
5	0.0017*
6	0.3133
7	0.1607
8	0.1257
9	0.9806
10	0.1806

*Significant at .05 alpha level

Table 7

Mean Ratings for Professors, Associate Professors,
and Assistant Professors on 10 Evaluation Items

<u>Rating No.</u>	<u>Profs.</u>	<u>Asso. Profs.</u>	<u>Asst. Profs.</u>
1	1.87	1.63	1.72
2	2.21	1.95	1.93
3	2.16	1.84	1.96
4	2.43	2.06	2.07
5	2.70	2.33	2.28
6	3.10	3.21	3.24
7	3.48	3.67	3.65
8	2.45	2.37	2.27
9	2.14	1.93	2.03
10	2.55	2.09	2.26
	N = 59	N = 75	N = 130

$\underline{F}$ (10,122) = 2.82225, $\underline{p} > 0.0038$. The univariate ANOVA indicated the evaluation items on which the two groups' ratings differed significantly (see Table 8).

Hotelling's $\underline{T}^2$ test indicated a difference between the ratings of associate professors and assistant professors at the .05 level of significance, $\underline{F}$ (10,192) = 2.31436, $\underline{p} > 0.0136$. The univariate analysis was used as a guide to the ratings on which the means for the two groups differed significantly (see Table 9).

The tests which were applied to the data were significant factors in determining whether the null hypothesis should be accepted or rejected. The summary measures were directed toward specific questions which this study sought to answer: were there student evaluation measures that were closely associated with tenure status, and did faculty members in one professorial rank receive better ratings than faculty members in another rank?

Table 8
Results of Univariate ANOVA for Professors
and Associate Professors

<u>Rating No.</u>	<u>Sign. Level</u>
1	0.0065*
2	0.0062*
3	0.0006*
4	0.0003*
5	0.0014*
6	0.3138
7	0.0538*
8	0.4826
9	0.0136*
10	0.0001*

*Significant at .05 alpha level

Table 9
Results of Univariate ANOVA for Associate
Professors and Assistant Professors

<u>Rating No.</u>	<u>Sign. Level</u>
1	0.2090
2	0.8288
3	0.1251
4	0.8651
5	0.5828
6	0.6698
7	0.8204
8	0.2016
9	0.1214
10	0.0571

CHAPTER V

ANALYSIS AND DISCUSSION OF DATA

The problem of this study was to examine the null hypothesis that there were no differences in the ratings received by tenured and nontenured faculty in student evaluations conducted in the College of Arts and Sciences at the University of Florida. This study further sought to determine if there were student evaluation measures that were closely associated with tenure status and if there were differences among the ratings of faculty members in the ranks of professor, associate professor, and assistant professor.

The 268 study participants represented 75% of the 355 College of Arts and Sciences faculty members. Teaching effectiveness was measured on the basis of the mean rating received by faculty members on each of the 10 items on the student evaluation form.

Faculty Members Classified by Group

On the basis of both their tenure status and professional rank, the study participants were classified into six groups: tenured professors, tenured associate professors, tenured assistant professors, nontenured associate

professors, nontenured assistant professors, and nontenured instructors. The largest of these six groups was that of the nontenured assistant professors, with 116 entries.

A multivariate one-way analysis of variance, used to test for differences among the six groups, showed that the ratings for the groups were different at the .05 level of significance. The univariate ANOVA was used as a guide to where these differences might be found among the groups. The means for items 2, 3, 4, 5, and 10 were significantly different at the .05 alpha level.

Simultaneous confidence intervals were set up for the group means on the five ratings that showed significant differences. The confidence intervals, however, did not indicate where significant differences were among the groups. What the confidence intervals did indicate was that no significant difference could be seen through an analysis of one rating at a time. This was further evidence that there was underlying interaction among the 10 variables on the evaluation instrument

Discriminant Analysis on Ten Measures

Data from the 10 items on the evaluation instrument were used to determine whether a discriminant analysis would make a distinction between tenured and nontenured faculty. In the discriminant analysis, the group means for the six groups—three of which were tenured and three of which were nontenured—on the 10 evaluation items were used to test

for group differences on 10 measures (the 10 item ratings). The group centroids, as determined by one group's maximal difference from another group on the 10 measures, served as one of the determinants in the reclassification of group members. The other determinant was the group member's ratings. Each faculty member was reclassified as a member of a group on the basis of the extent to which his ratings on the 10 evaluation items resembled those ratings of a group's members.

The discriminant analysis made a very definite distinction between tenured and nontenured faculty. In the regrouping process, only one tenured faculty member was classified as nontenured; and only one nontenured faculty member was classified as tenured. This indicated clearly that the student ratings for tenured and nontenured faculty members were different. Had there been no difference between the ratings of tenured and nontenured faculty, one-half of the tenured subjects would have been placed into nontenured groups; and one-half of the nontenured faculty would have been placed into tenured groups.

Other results of the regrouping process in the discriminant analysis were seen in the change in the number of cell entries. Approximately one-half, 30 out of 59, of the tenured professors were placed into the groups of tenured associate professors and tenured assistant professors, 15 entries being regrouped into each. Thus, tenured professors

were shown to be similar to tenured associate professors and to tenured assistant professors. One-third of the tenured associate professors were reclassified in equal numbers as tenured professors and tenured assistant professors. The one tenured person who was regrouped as a nontenured associate professor was a tenured associate professor. One-half of the 14 tenured assistant professors were reclassified, four as tenured professors and three as tenured associate professors.

Among the nontenured groups the most significant shifting of group members was found among the nontenured associate professors. Two-thirds of the group's members were reclassified in equal numbers as nontenured assistant professors and nontenured instructors. The single nontenured subject who was reclassified as a tenured associate professor was a nontenured associate professor. Approximately one-fourth of the 116 nontenured assistant professors were reclassified, 25 as nontenured instructors and 7 as nontenured associate professors. Only one of the four nontenured instructors was shifted; he was placed into the group of nontenured assistant professors.

Of the total 100 out of 268 subjects who were placed into different groups through the discriminant analysis, 60 were tenured faculty members and 40 were nontenured faculty members. Thus, the percentage of incorrectly grouped faculty members was greater for tenured faculty members than for nontenured faculty: 43% as compared to 31%.

Comparison of Student Ratings for Tenured
and Nontenured Faculty

Because significant group differences were shown through the multivariate analysis and the discriminant analysis for groups identified by both their tenure status and rank, comparisons were made to find out where significant differences were between groups identified by their tenure status only and between groups identified by rank only. Comparisons of interest were those between (a) tenured and nontenured faculty, (b) professors and associate professors, and (c) associate professors and assistant professors.

Hotelling's T^2 test was used to test for differences between the ratings for tenured and nontenured faculty. Hotelling's T^2 test, an extension of the t test to 10 dimensions, is identical to the usual t^2 ; which would have been used if there had been one rating item instead of 10 items. This test indicated that the ratings for tenured and nontenured faculty members were different at the .05 level of significance.

The univariate ANOVA showed that the ratings for the two groups were significantly different at the .05 alpha level on three of the evaluation items: 2, 4, and 5. Table 5 shows that nontenured faculty members received better ratings on all three of these items. Although the remaining differences were not significant, nontenured faculty received better ratings on items 3, 6, 7, 8, and 10 (a high

rating of 5-4 on items 6 and 7 was positive, whereas a low rating of 1-2 on all of the other items was positive). On items 1 and 9, the ratings for the two groups were the same. Tenured faculty did not receive better ratings than nontenured faculty on any of the items.

Comparison of Student Ratings for Professors and Associate Professors

Hotelling's T^2 test was also used to test for differences between the student ratings for professors and associate professors. The ratings were different at the .05 level of significance. The univariate ANOVA indicated that on all of the items except 6 and 8 the ratings were significantly different. Although two of the rating differences for the two groups were not significant, associate professors received better ratings than professors on all 10 evaluation items (see Table 7). The surface differences between the ratings for these two groups were quite distinct.

Because one-fourth of the tenured professors had been reclassified as tenured associate professors in the discriminant analysis of groups identified by tenure status and professorial rank, it had been assumed that these two groups were quite similar. The rating differences shown in the subsequent analysis—by rank only—of these two groups implied, however, that the 15 members who had been shifted from the group of tenured professors to the tenured associate professors' group were those professors who had the

best ratings among the group of professors. This seemed logical when it was observed that the mean ratings for associate professors on the evaluation items were slightly better than those for assistant professors. For this reason the professors with the best ratings should have been grouped with the tenured associate professors instead of the tenured assistant professors. This seemingly logical view did not necessarily hold true when it was observed that in the mean ratings for six groups (see Table 2) the ratings for tenured assistant professors were slightly better than those for tenured associate professors on items 2, 5, 6, 8, and 9. The mean ratings for the two groups were the same on items 1 and 4. What the analysis of the data did show was that in the discriminant analysis one-half of the 30 tenured professors who were reclassified had been placed into the tenured associate professors' group and one-half had been placed into the tenured assistant professors' group. This equal distributing of group members between the two groups indicated that the ratings for tenured associate professors and tenured assistant professors were the same. And for this reason, the group which professors became members of made no difference.

Thus, the analysis of faculty members by rank only pointed out rating differences that could not be seen when tenure status and professorial rank were treated as one variable. On the basis of rank only, it was clear that the

ratings for associate professors were better than those for professors.

Comparison of Student Ratings for Associate Professors and Assistant Professors

A comparison of the student ratings for associate professors and assistant professors, using Hotelling's T^2 test, indicated that the ratings were significantly different at the .05 alpha level. But the univariate ANOVA did not show that the mean ratings for the two groups on any of the items were significantly different. The only rating that could have been regarded as significantly different was that for item 10, which had a significance level of 0.0571.

Table 7 does show, however, that the mean ratings for associate professors on items 1, 3, 9, and 10 were slightly better than those for assistant professors. Assistant professors received slightly better ratings on items 5, 6, and 8. Ratings for the two groups were the same on items 2, 4, and 7.

Summary Statement

The analysis of the data indicated that there were significant differences in the student ratings of teaching effectiveness for (a) tenured and nontenured faculty, (b) professors and associate professors, and (c) associate professors and assistant professors. The three sets of rating comparisons showed (a) that nontenured faculty members received better ratings than tenured faculty,

(b) that associate professors received better ratings than professors, and (c) that associate professors received slightly better ratings than assistant professors.

In regard to the question of whether there were student evaluation measures that were closely associated with tenure status, the analysis of the data did not indicate that there were definite patterns of difference by group on the evaluation items. Measures 2, 4, and 5, however, did point out significant differences in the ratings received by tenured and nontenured faculty members.

CHAPTER VI

FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The focus of this study was upon student evaluations and the tenure status of faculty members. The purpose of the study was to examine the null hypothesis that there were no differences in the ratings received by tenured and non-tenured faculty in student evaluations conducted in the College of Arts and Sciences at the University of Florida.

Findings

The results of this investigation did not substantiate the null hypothesis. The ratings of tenured and nontenured faculty were found to be significantly different at the .05 alpha level. Significant differences were also found in the ratings received by faculty members classified according to professorial rank: professor, associate professor, and assistant professor.

The data were analyzed through the use of multivariate analysis of variance, analysis of variance, a discriminant analysis, and Hotelling's T^2 test. The multivariate analysis was the first test applied to the data; the null hypothesis of this study was rejected on the basis of the results of this test procedure. Simultaneous confidence intervals

were then set up to determine which group means differed from their hypothesized values and were thereby contributing to the rejection of the overall hypothesis. These groups were not identified when the confidence intervals were established.

Through the test procedures of the multivariate analysis of variance and the confidence intervals, it was determined that the rating items on the student evaluation instrument were correlated variables. As such, no rating item—used alone—would make a meaningful distinction between groups of faculty members.

The univariate analysis of variance was used only as a guide to the item ratings on which group means differed significantly. It was not used as a final test measure.

The results of the discriminant analysis showed that tenured and nontenured faculty members formed two distinct groups. The two groups differed to the extent that when tested to see if there was enough similarity between the groups' ratings to cause group members to move from one tenure category into the other, the groups did not exchange members. The discriminant analysis also showed that in terms of an individual's distance from the centroid of his respective group—as defined by both tenure status and faculty rank—37% of the faculty members were incorrectly classified.

Hotelling's T^2 test, which was used to test for differences between the ratings of tenured and nontenured

faculty, professors and associate professors, and associate professors and assistant professors, indicated that there were significant differences at the .05 alpha level in all three sets of comparisons. A study of the group means for each evaluation item and the univariate analysis of variance showed (a) that nontenured faculty members received better ratings than tenured faculty on eight of the 10 evaluation items—with three of these rating differences being significant at the .05 alpha level; (b) that associate professors received better ratings than professors on all 10 evaluation items—eight of which were significantly different at the .05 alpha level; and (c) that associate professors received better ratings than assistant professors on four of the evaluation items and assistant professors received better ratings than associate professors on three of the 10 items—none of the item differences being significant at the .05 alpha level.

Because the multivariate analysis of variance indicated that the items on the student evaluation instrument were correlated variables, the differences—though significant at the .05 alpha level—between tenured and nontenured faculty, professors and associate professors, and associate professors and assistant professors are more accurately reported in terms of the number of positive ratings which a group received.

Conclusions

The findings of this study support the following conclusions:

1. Tenure is negatively related to faculty members' teaching effectiveness as measured by student evaluations.
2. The role of student evaluations is apparently one of little importance to the tenure status of faculty members.
3. According to student evaluations, associate professors are better teachers than both professors and assistant professors.

Recommendations

The results of this study alone cannot provide definitive justification for the use or of the value of student evaluation of teaching effectiveness in tenure considerations. It is, therefore, necessary for additional study to be given to the factors examined in this investigation.

It is recommended that attention and further study be given to the following relevant concerns:

1. The teaching effectiveness of faculty members prior to and following the granting of tenure;
2. The granting of tenure for a specified period, with independent evaluations for tenure at the end of each set time period;

3. The weight attached to teaching effectiveness in tenure considerations;

4. The role of student evaluation of teaching effectiveness in tenure considerations.

BIBLIOGRAPHY

- Astin, Alexander W. and Lee, Calvin B. T. "Current Practices in the Evaluation and Training of College Teachers." In Improving College Teaching, pp. 296-311. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.
- Barr, Anthony J. and Goodnight, James H. A User's Guide to the Statistical Analysis System. Raleigh: Sparks Press, 1972.
- Bauer, Roger D. "Security Versus Freedom: Some Faculty Views on the Tenure Process." Phi Delta Kappan 54 (December 1972):279-80.
- Blai, B. "Faculty-Ranked Importance of Several Criteria in Pay-Promotion Matters." Scientia Paedagogica Experimentalis 9 (1972):5-17.
- Budig, Gene A. and Decker, C. Richard. "Assistant Professors, Disillusioned by Tenure, Believe Unionization Offers Better Protection." Phi Delta Kappan 55 (October 1973): 143-44.
- Centra, John A. "Self-Ratings of College Teachers: A Comparison with Student Ratings." Journal of Educational Measurement 10 (Winter 1973):287-95.
- Costin, Frank. "A Graduate Course in the Teaching of Psychology: Description and Evaluation." Journal of Teacher Education 19 (Winter 1968):425-32.
- Costin, Frank; Greenough, William T.; and Menges, Robert J. "Student Ratings of College Teaching: Reliability, Validity, and Usefulness." Review of Educational Research 41 (December 1971):511-35.
- Crawford, P. L. and Bradshaw, H. L. "Perception of Characteristics of Effective University Teachers: A Scaling Analysis." Educational and Psychological Measurement 28 (Winter 1968):1079-85.
- Downie, N. W. "Student Evaluation of Faculty." Journal of Higher Education 23 (December 1952):495-96.

Drucker, A. J. and Remmers, H. H. "Do Alumni and Students Differ in Their Attitudes Toward Instructors?" Journal of Educational Psychology 42 (March 1951):129-43.

Eble, Kenneth E. Professors as Teachers. San Francisco: Jossey-Bass, 1972.

_____. The Recognition and Evaluation of Teaching. Salt Lake City: Project to Improve College Teaching, 1970.

Edmonson, J. B. and Mulder, F. J. "Size of Class as a Factor in University Instruction." Journal of Educational Research 9 (January 1924):1-12.

"Elimination of Faculty Tenure." Intellect 101 (March 1973): 346.

El-Khawas, Elaine H. and Furniss, W. Todd. Faculty Tenure and Contract Systems: 1972 and 1974. Washington, D. C.: American Council on Education.

Freeman, Thomas M. and Rossmeier, Joseph G. A New Look at Tenure: A Management Imperative. East Lansing, Mich.: ERIC Document Reproduction Service, ED 080 097.

Gage, N. L. "The Appraisal of College Teaching." Journal of Higher Education 32 (January 1961):17-22.

Gustad, John W. "Evaluation of Teaching Performance: Issues and Possibilities." In Improving College Teaching, pp. 265-81. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.

Guthrie, E. R. "The Evaluation of Teaching." Educational Record 30 (January 1949):109-15.

_____. The Evaluation of Teaching: A Progress Report. Seattle: University of Washington, 1954.

Heilman, J. D. and Armentrout, W. D. "The Rating of College Teachers on Ten Traits by Their Students." Journal of Educational Psychology 27 (March 1936):197-216.

Howe, Harold. "Less Teaching, More Conversation." In Improving College Teaching, pp. 259-64. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.

Hughes, Richard E. "How to Solve the Tenure Problem: Departmental or College Professorship." College Management 9 (January 1974):7, 27.

- Johnson, James A. "Instruction: From the Consumer's View." In Improving College Teaching, pp. 288-92. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.
- Kent, Laura. "Student Evaluation of Teaching." In Improving College Teaching, pp. 312-43. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.
- Kramer, Clyde Y. A First Course in Methods of Multivariate Analysis. Blacksburg, Va.: By the Author, 1972.
- Lang, Theodore H. "Teacher Tenure as a Management Problem." Phi Delta Kappan 56 (March 1975):459-62.
- Lee, C. B. T., ed. Improving College Teaching. Washington, D. C.: American Council on Education, 1967.
- Lieberman, Myron. "Tenure: A New High-Priority Issue." Phi Delta Kappan 56 (March 1975):450.
- Lovell, G. D. and Haner, C. F. "Forced-Choice Applied to College Faculty Rating." Educational and Psychological Measurement 15 (Autumn 1955):291-304.
- McKeachie, Wilbert J. Teaching Tips: A Guide for the Beginning Teacher. 6th ed. Lexington, Mass.: D. C. Heath and Co., 1969.
- _____. "Research in Teaching: The Gap Between Theory and Practice." In Improving College Teaching, pp. 211-39. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.
- McKeachie, Wilbert J.; Lin, Yi-Guang; and Mann, William. "Student Ratings of Teacher Effectiveness: Validity Studies." American Educational Research Journal 8 (May 1971):435-45.
- McKeachie, Wilbert J. and Solomon, D. "Student Ratings of Instructors: A Validity Study." Journal of Educational Research 51 (January 1958):379-82.
- Magarrell, Jack. "Trying to Devise a Way to Measure the Benefits of Higher Education." Chronicle of Higher Education, 1 April 1974, p. 12.
- Mann, William R. "Is the Tenure Controversy a Red Herring?" Journal of Higher Education 44 (February 1973):85-94.

- Maslow, A. H. and Zimmerman, W. "College Teaching Ability, Scholarly Activity and Personality." Journal of Educational Psychology 47 (March 1956):185-89.
- Metzger, Walter P. "Academic Tenure in America: A Historical Essay." In Faculty Tenure, pp. 93-159. Edited by W. R. Keast. San Francisco: Jossey-Bass, 1973.
- Musella, D. and Rusch, R. "Student Opinion on College Teaching." Improving College and University Teaching 16 (Spring 1968):137-40.
- Remmers, H. H. and Elliott, D. N. "The Indiana College and University Staff-Evaluation Program." School and Society 70 (July-December 1949):168-71.
- Rohrer, J. H. "Large and Small Sections in College Classes." Journal of Higher Education 28 (May 1957):279-81.
- Rollo, Mike, ed. Course and Teacher Evaluation Handbook. Gainesville: University of Florida, 1974.
- Russell, H. E. and Bendig, A. W. "Investigation of the Relations of Student Ratings of Psychology Instructors to Their Course Achievement When Academic Aptitude Is Controlled." Educational and Psychological Measurement 13 (1953):626-35.
- Samalonis, Bernice. "Ratings by Students." Improving College and University Teaching 15 (Winter 1967):11.
- Scully, Malcolm G. "Improve Teaching, Prevent Stagnation, Group Urges." Chronicle of Higher Education, 18 March 1974, p. 1.
- Semas, Philip W. "Faculty Firings Soar as Slump Hits Enrollment." Chronicle of Higher Education, 28 January 1974, pp. 1, 3.
- Shulman, Carol H. "The Tenure Debate." In Research Currents. Washington, D. C.: ERIC Document Reproduction Service, ED 054 391.
- Smith, Bardwell, ed. The Tenure Debate. San Francisco: Jossey-Bass, 1973.
- Stewart, C. T. and Malpass, L. F. "Estimates of Achievement and Ratings of Instructors." Journal of Educational Research 59 (April 1966):347-50.

Tatsuoka, Maurice M. Discriminant Analysis: The Study of Group Differences. Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

_____. Multivariate Analysis: Techniques for Educational and Psychological Research. New York: Wiley & Sons, Inc., 1971.

Tolor, Alexander. "Evaluation of Perceived Teacher Effectiveness." Journal of Educational Psychology 64 (February 1973):98-104.

University of Florida. Constitution (1972).

Voeks, V. W. and French, G. M. "Are Student-Ratings of Teachers Affected by Grades?" Journal of Higher Education 31 (June 1960):330-34.

Wallace, Trevette, ed. Course and Teacher Evaluation Handbook. Gainesville: University of Florida, 1973.

Whitfield, Raymond P. and Brammer, Lawrence M. "The Ills of College Teaching: Diagnosis and Prescription." Journal of Higher Education 44 (January 1973):1-14.

Wilson, Logan. "The Professor and His Roles." In Improving College Teaching, pp. 99-109. Edited by C. B. T. Lee. Washington, D. C.: American Council on Education, 1967.

Wood, Kenneth; Linsky, Arnold S.; and Straus, Murray A. "Class Size and Student Evaluations of Faculty." Journal of Higher Education 45 (October 1974):524-34.

BIOGRAPHICAL SKETCH

Lily M. Loretta James was born in Clearwater, Florida, on October 27, 1941. She attended schools in Clearwater, graduating from Pinellas High School in 1959. After moving, along with her family, to St. Petersburg, Florida, in 1959; she enrolled at Gibbs Junior College. She received the Associate of Arts degree with honors from Gibbs in 1961. From 1962-64 she attended Florida A. and M. University, where she was awarded the Bachelor of Arts degree in English with distinction in 1964.

Miss James taught English in St. Petersburg for five years and returned to school as a full-time student in 1969. In 1967 she was awarded a grant by Northeastern Conference of Seventh-day Adventists for summer study at Atlantic Union College. Miss James received the Master of Arts degree in English (language) from Indiana University in 1970.

After graduation, she taught English at Oakwood College for one year, returning to teach in Pinellas County in 1971. In the summer of 1972, Miss James studied at the University of South Florida. She was awarded an SDI Fellowship for doctoral study at the University of Florida in 1973.

Miss James is a member of Phi Delta Kappa, Pi Lambda Theta, the American Association for Higher Education, the American Association of School Administrators, and the National Council of Teachers of English.

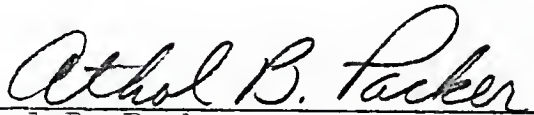
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Kern Alexander, Chairman
Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

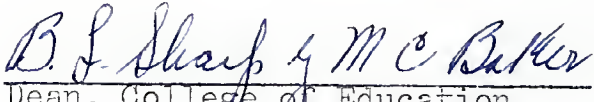

James L. Wattenbarger
Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Athol B. Packer
Associate Professor of Education

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

June, 1975


Dean, College of Education

Dean, Graduate School

